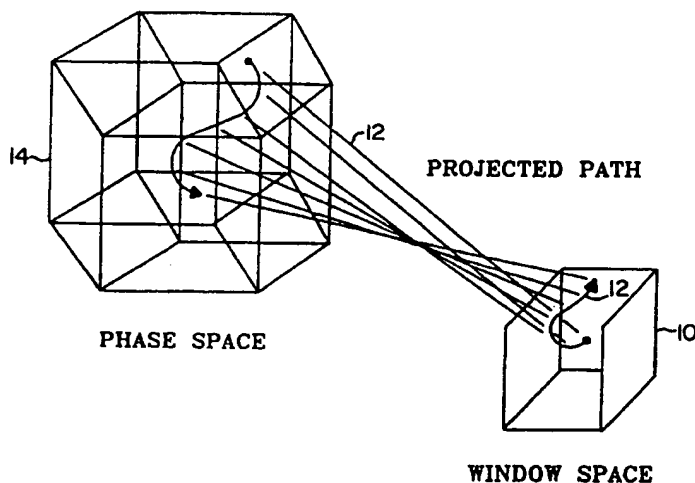




## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> : <b>G10L 5/02</b>	<b>A1</b>	(11) International Publication Number: <b>WO 99/13455</b> (43) International Publication Date: 18 March 1999 (18.03.99)
(21) International Application Number: PCT/US98/18290 (22) International Filing Date: 3 September 1998 (03.09.98) (30) Priority Data: 08/924,179                5 September 1997 (05.09.97)    US (71) Applicant: THE BOARD OF TRUSTEES OF THE UNIVERSITY OF ILLINOIS [US/US]; 352 Henry Administration Building, 506 South Wright Street, Urbana, IL 61801 (US). (72) Inventors: BARGAR, Robin; 2019 Burlison Drive, Urbana, IL 61801 (US). CHOI, Insook; 2019 Burlison Drive, Urbana, IL 61801 (US). GOUDESEUNE, Camille, M.; 603 West Nevada Street #1, Urbana, IL 61801 (US). (74) Agent: FALLON, Steven, P.; Greer, Burns & Crain, Ltd., Sears Tower, Suite 8660, 233 South Wacker Drive, Chicago, IL 60606 (US).	(81) Designated States: AU, CA, JP, KR, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).  <b>Published</b> <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>	

(54) Title: SYSTEM AND METHOD FOR INTERFACING SOUND SYNTHESIS TO MOVEMENT



## (57) Abstract

High-dimensional descriptions of data or "manifolds" (40, 62), such as the plurality of parameters that characterize sound, are mapped into two or three dimensions. A "manifold interface" provides the translation to sound signal generation (62) or to other multidimensional data bases. The manifold interface technology translates movements (30) performed in the three-dimensional space into movements in a high-dimensional parameter space. A three-dimensional virtuality reality environment (20) captures movement (30) of an object. The system maps (12) the location of the object to a location in a multidimensional phase space (14). The phase space (14) represents all the permissible combinations of parameter values of an algorithm where trajectories (30) of input gestures are encoded. For display, the systems reduces (12) the dimension of the phase space (56) to three dimensions (50) of a window space (10), which translates the multidimensional information for the phase space to render an image (50) of relationships between the location of the object and the numerical values assumed by the parameters in correspondence to the location of the object.

*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Larvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
BF	Burkina Faso	GR	Greece		Republic of Macedonia	TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakistan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

## SYSTEM AND METHOD FOR INTERFACING SOUND SYNTHESIS TO MOVEMENT

### TECHNICAL FIELD

The invention relates generally to electronic sound synthesis. It specifically  
5 relates to interfacing sound synthesis with movement, including rendering and reproduction.

### BACKGROUND ART

Sounds as an auditory percept are little understood in terms of their potential  
to play a significant role for research and creative projects particularly involving high-  
dimensional systems. The applications of sound in interface design include voice  
10 recognition, teleconferencing, audio archiving, sound localization, audio alarms, audio cues,  
carcons, and data sonifications. Many of these applications serve the purpose of enhancing  
visualization or compensating for visual overload. For example, audio cues are sounds for  
location identification guiding visual search for a point of interest. Among these applications  
data sonification comes close to utilizing auditory percepts for enhancing an understanding  
15 of data.

To bring auditory percepts into research projects involves (1) designing sounds  
for an optimal representation of systems' behaviors, and (2) incorporating sounds in  
interactivity. For exploring systems, observers often encounter cumbersome tasks such as  
entering control data by typing or creating input files. Output data are also observed often  
20 in the form of numbers or graphic representations. When exploring high-dimensional  
systems a need exists for alternative ways of interacting with the systems. An efficient  
method for entering control data with real-time observation of the consequences are keys to  
an intuitive exploration. The use of sounds has been observed to offer efficient and  
perceptive learning in massive parameter space. As a system output, sound functions as an  
25 auditory feedback, linking full circle in an exploration process for observers to monitor their

own interaction as well as the behavioral changes of systems under study.

The unique characteristics of sound lies in the omnidirectional characteristic of acoustic signals. This characteristic can be understood in two ways. First, the obvious meaning of "omnidirectional" refers to the way sounds propagate in space. This accounts  
5 for the physics of sounds such as diffusion, reflection, and diffraction as well as our perceptual ability to process the spatial distribution of sounds. Secondly, the term "omnidirectional" can be understood from a compositional point of view focusing on acoustic materials or elements, their pitch and rhythmic relationships, their sizes in units and groups. In other words, we can also apply "omnidirectional" to refer to classes of sounds  
10 within a *material differentiation space*. By listening to the way classes of materials are derived from an original set and developed through or without transitional states, one achieves a dynamical observation. An example can be found where the acoustic material differentiation is based upon the content area of an "information space".

The prior art lacks the advantage of a system and method for representing the  
15 omnidimensional sound characteristics in a comprehensible manner. Such a system and method would translate multi-dimensional sound representations or "manifolds" into two or three dimensions that can be understood and manipulated more readily by the observer. A "manifold interface" provides such a translation or mapping, which can further be applied to other multi-dimensional control systems in addition to sound signal generation and sound  
20 analysis. The benefit of this mapping is that humans understand movement in 2D and 3D, whereas it may be difficult to intuitively grasp high-dimensional spaces. The manifold interface technology translates movements performed in the intuitive space into movements in a space that is otherwise difficult to grasp intuitively.

## DISCLOSURE OF THE INVENTION

25 In view of the above, a system and method is provided for navigating a manifold in a high-dimensional space. According to the system of the invention, a three-dimensional virtual reality environment operates to capture movement of an object in relation to the manifold. The captured movement is coupled to be communicated to means for generating a phase space representing at least one dimension of the manifold in relation to  
30 location of the object. Means are also provided to generate a window space for translating

information from the phase space to render an image of the relation between the location of the object and the manifold. The image rendered is capable of being displayed.

According to the method of the invention, movement of an object in relation to a manifold is captured in a three-dimensional virtual reality environment. The captured movement is then communicated in order to generate a phase space representing at least one dimension of the manifold in relation to the location of the object. A window space is also generated for translating information from the phase space to render an image of the relationship between the location of the object and the manifold. The image can then be displayed.

In one preferred embodiment of the invention, the manifold is a set of numerical values to be applied to parameters controlling an audio signal generator and the object is the movement of a composer. In this configuration, a composer's movements can be translated to the generation or composition of sounds and music. Alternative embodiments are contemplated where the system and method can be used to plan robotic movement or compose movement of animated figures. Representations of complex multi-dimensional systems can also be displayed in a two or three dimensional manner. Such systems include, for example, particle systems or gaseous and liquid volumes and fields. In another embodiment, the manifold interface can be applied to enable the intuitive control of these systems.

The system and method of the invention provides the advantage of a manifold interface between high-dimensional systems and a two or three dimensional representation thereof. The translation of the high-dimensional manifold to a two or three dimensional image allows for easy and more readily understandable applications. The manifold interface thus allows control over a reasonable sub-set of all of the points of the manifold without a need to access all of the points at one time.

These and other features and advantages of the invention will become apparent upon a review of the following detailed description of the presently preferred embodiments of the invention taken in conjunction with the appended drawing.

#### BRIEF DESCRIPTION OF THE DRAWINGS

FIGURE 1 shows an example of embedding in a window space;

FIG. 2 provides a bijective map between phase space and window space;

FIG. 3 illustrates a view of a path in a window space;

FIG. 4 illustrates a view of a surface in a window space;

FIG. 5 is a flow chart showing the control flow of the preferred manifold  
5 controller environment;

FIG. 6 shows the vowel sounds created by a path controlling CHANT, a  
software-based audio signal generator;

FIG. 7 shows a software structure for mapping between component spaces in  
sound rendering according to the invention and performed by the manifold controller.

## 10 BEST MODE FOR CARRYING OUT THE INVENTION

In the most general description, the manifold interface is a system and method  
for mapping a high-dimensional space of a computational model into a 2- or a 3-dimensional  
space, where signals are generated by an observer's movement controlling a computer input  
device (not shown). As those skilled in the art will appreciate, such devices may include a  
15 mouse, joystick or 3D position sensor to name a few. The 2- or 3-D input signals are mapped  
back into the high-dimensional space to indicate positions and position changes in the high-D  
space. Therefor the manifold interface allows intuitive movements in 2- or 3-dimensions to  
be equated with changes in a high-dimensional space which is not intuitive. In one presently  
preferred embodiment, the manifold interface system and method is applied to sound  
20 synthesis. As those skilled in the art will further appreciate, other applications of the  
manifold interface are contemplated without departing from the essential spirit and scope of  
the invention. Presently contemplated alternate applications are described in detail below.

The term "window space" is used herein to refer to the 2- or 3-D space  
described above. The term "phase space" is used herein to refer to the high-dimensional  
25 space of the computational model. For the purposes of this description, "high-dimensional"  
means four or more dimensions. The manifold interface has been successfully applied to  
mappings of up to 30 or 40 dimensions, and there is theoretically no upper limit. Changing  
positions in a high-dimensional space is sometimes referred to herein as "navigating" a high-  
dimensional space. This term also refers to the common scenario where the observer is  
30 looking for features in the space. These terms are discussed in greater detail below.

In the presently preferred embodiment of the invention, a manifold controller (MC) is a set of C++ classes linking graphics, hardware input devices, and sound synthesis engines. The MC can be defined as an interactive graphical sound generation tool and composition interface involving computational models; computational models may be sound synthesis models, composition algorithms, or any other numerical models such as chaotic systems. Its application is preferably scalable from immersive virtual environments to desktop workstations.

The manifold interface provides graphical lines and surfaces as an interface to manifolds of greater than three dimensions. The interface allows a user to navigate in a high-dimensional parametric space from a visual display having a continuous gesture input system with at least two degrees of freedom. The preferred embodiment includes 3D gesture input and 3D display. For workstations supporting 2D controllers and 2D graphical display the references can be scaled down.

The graphic interface is also preferably linked to the NCSA Sound Server, which provides a real-time sound synthesis and algorithmic composition environment as well as non-real-time synthesis for demanding computational tasks. The architecture of the Sound Server allows the interface to be used concurrently for low-level control of synthesis engines and high-level control of composition parameters. The Sound Server is located on the campus of the University of Illinois in Urbana, Illinois, United States of America.

## ORGANIZATION AND REPRESENTATION OF CONTROL PARAMETER SPACE

Control parameters and all combinations of them involving computational models such as numerical models of physical forces and their interactions present a massive space to explore. The MC seeks for efficient system access by organizing control parameters so that one can easily manipulate them into different combinations with rapid annotation capabilities to keep track of sequences of actions. It is also preferable that the representation of the systems have visual simplicity while maintaining an accuracy of its relationship to the actual states of the systems. This visual simplicity is an important factor to engage observers in an intuitive exploration.

In organization and representation of control parameter space three spaces are

distinguished: control space, phase space and window space. The term "control space" is used on a conceptual basis to implicitly refer to both phase and window space as a couple, whereas the terms "phase space" and "window space" have special meanings in terms of technical relationships. The *phase space* of a system means the traditional n-dimensional Euclidean space where points --n-tuples of real numbers -- correspond to states of a parameterized system. The phase space represents all the permissible combinations of parameter values of an algorithm where trajectories of input gestures are encoded. A literal representation of high-dimensional phase space will be visually undifferentiable resulting in the loss of orientation. Thus, a representation space is needed with data reduction from arbitrary high-dimensional phase space to 3D space in perceptible form.

This represented phase space is referred to herein as a *window space*. Referring to Figure 1, the window space 10 defines how a three-dimensional visual representation 12 is embedded in the high-dimensional phase space 14. A three-dimensional visual display 12 is used as a window onto the manifold so that an observer inputs changes to the system through the window space 10. An observer may effectively control the window space 10 by panning and zooming in phase space 14.

In the presently preferred embodiment, the window is displayed in the CAVE environment or a similar 3D view. The CAVE environment is located at the University of Illinois in Urbana, Illinois, United States of America. In one alternate embodiment, the cursor is depicted on a 2D graphical surface using a standard computer workstation display (not shown) and responds to 2D gesture-input devices such as a mouse (not shown), and to voice and keyboard (not shown) commands.

The manifold interface provides reference points to calibrate certain points in the window space 10 with points in the phase space 14. Reference points are referred to as "generating points". These are so-named because the mapping begins process by specifying reference points which are used to generate a mapping. The manifold interface technology provides smooth mapping of points between the generating points. The total set of high-dimensional points connecting and passing through all of the generating points forms a surface in the high-dimensional space, a subset of all the points in that space. This surface is referred to in mathematics as a "manifold." In the preferred embodiment, the MC allows "generating points" to be positioned by the observer at specific 2D or 3D points in the



window space 10. A genetic algorithm GA is applied to find the smooth connections between generating points. The GA then finds the smoothest possible manifold between these points.

5 A window space 10 provides a domain for generating and modifying classes of control point sets. These points represent combinations of parameter values as user-specified, and they are associated with particular sounds. This association of the sounds in conjunction with positional orientation in window space 10 enhances the ability to identify boundaries where character shifts occur in states of the system.

10 Preferably, the system is able to visit the generating points and move smoothly between them. Since phase space 14 may involve twists and bends during the embedding process, the embedding should be continuous and "simple" while preserving a maximum amount of information. For data reduction from phase space 14 to window space 10, the GA is employed to find a near-optimal window space 10 by starting with a random population of possible solutions and allowing the better solutions to "multiply" and create offspring. For  
15 maximal representation of the structure in a phase space 14, in the region of generating points, a fitness function and bit-representation of a solution is applied as the system interprets "structure" as the matrix of Euclidean distances between points. The states the GA explores are sets of points in the window space 10, represented as vectors of fixed-point numbers; the fitness function measures the error between the original distance matrix and the  
20 matrix for a particular set of points in window space 10.

A good definition of the window space 10 is critical to all aspects of the Manifold Controller. There is an inevitable information loss as dimensions are reduced. The nature of the information loss affects the size and shape of the manifold region that can be controlled from a window space 10. As this problem is difficult and impossible to solve  
25 precisely (at least when the phase space 14 has more than 3 dimensions), a GA and fitness function help to solve such a complex problem.

The image of the generating points in the window space is extended to a 3-dimensional lattice where lines through the generating points are more or less parallel to the principal axes of the space. All points in the lattice are then used in a reversal of the previous  
30 GA to produce a corresponding lattice of similar geometry in the phase space 14. Referring to Figure 2, to map one point in the window space 10 to the one in phase space 14, first the

lattice cell 20 where the point belongs has to be searched. Then its coordinates in the cell are found based on a tetrahedral decomposition 22, 24 of the cell 20. The corresponding cell 26 and coordinates in the phase space 28 define the resultant point in the phase space 14. The inverse map is computed similarly. As a point's cell-coordinates exist and are unique under certain conditions which the cells satisfy (convexity, noncoincidence of corner vertices), this map from one space to cell-coordinates and back to another space exists and is bijective (see Figure 2). As the map is a patch of linear functions continuously connected, it is continuous as well.

To smooth out the map's nondifferentiable "edges," high-dimensional splines, preferably cubic B-spline volumes built on a perturbation of the 3-dimensional lattice in the product of the phase and window spaces, are used. In a Euclidean space, given a sequence of control points  $\{p_0, \dots, p_n\}$  and an index parameter  $u$ , the equation:

$$p(u) = \sum_{k=0}^n p_k N_{k,c}(u)$$

defines the B-spline curve for these control points, where  $N_{k,c}$  are the standard B-spline blending functions, i.e. polynomials of degree  $t-1$ . Cubic splines are preferred, hence  $t=4$ . Given a 3-dimensional lattice  $\{p_{j,k,l}\}$  of control points, its associated B-spline volume is defined by the equation:

$$p(u, v, w) = \sum_{i=0}^{n_1} \sum_{k=0}^{n_2} \sum_{l=0}^{n_1} p_{j,k,l} N_{j,4}(u) N_{k,4}(v) N_{l,4}(w)$$

over the index parameters  $u, v, w$ . Since generating points should map onto their images in the window space, the original lattice is perturbed in the product of the phase and window spaces with another GA to find a lattice whose use as a set of control points for a B-spline volume will yield this exact mapping. This search takes a long time to compute, because the GA's fitness function evaluates this spline equation for many values. The inverse computation is slower still, that of finding index parameter  $u, v, w$  which correspond to a

given point in the product space (equivalently, in one of its two component spaces). However, once these indices are found they provide the mapping between the component spaces without any linearizing steps such as the tetrahedral decomposition of a lattice cell.

The preferred software structure for the mapping is illustrated in FIG. 7. The  
5 silent application (client) 100 interacts with the sound authoring structure (SAS) 340 in three different ways. One interaction occurs during initialization. Afterwards, two kinds of message tags 324 are sent to the SAS, namely tags whose data arrays contain data information (affecting parameters of the sound algorithms 130 and renderers 90), and tags  
10 whose data arrays contain control information (affecting which algorithms 130 and renderers 90 are used and how they are connected).

The SAS has been implemented as a C-language library. Its structure can be considered a dynamic "snapshot" of the text in the audio description file. This file is chosen by client 100. Its name is passed in a message tag 324 to the tag translator 348. Upon receipt of this tag, the translator 348 reads and parses the statements in the audio description file  
15 (e.g., with a conventional Unix lexical analyzer and parser based on the tools "lex" and "yacc"), and based on that parsing constructs the list of message groups 360 with their corresponding rendering formats 364, as well as the run-time structures of the intelligent mixing/rendering entity 350. It may also pass control messages 380 to set up the algorithms 130 and renderers 90.

20 Specifically, after the initialization phase, during the running of the application 10, control information is passed straight through the SAS 340 from message tags 324 to control messages 380 and on to the algorithms 130 and renderers 90. Data information, on the other hand, is further processed before passing on as data messages 380 to the algorithms 130 and renderers. A message tag 324 (containing a key and a copy of the data array 322)  
25 from the application 100 is received by the tag translator. The tag translator 348 extracts the key and uses it to index into the table of message groups 360, for example, message group 360b in FIG. 7. Message group 360b points to a particular rendering format structure 364 by matching the tag 372 in the structure 364, while other groups 360x point to their own rendering format structures in the same way. The tag translator 348 sends a copy of the data  
30 array 322 which was in the message tag to each element in the message list 376. The elements of the message list 376, which may be referred to as "potential messages" or

"message patterns", are realized into actual data messages 380 by the act of giving them an actual data array 322. A final step occurs before the data messages 380 are sent on to the algorithms 130 and renderers 90: the data messages 380 are passed through the intelligent mixing and rendering entity 350, whose elements were set up during the initialization phase.

5

## PATHS AND DATA MANAGEMENT FEATURES

Using a hardware input device such as a wand (not shown), which permits three degrees of freedom in VR, by making arm gestures an observer may draw traces in window space 10. Examples of such traces 30 are illustrated in Figure 3. These traces 30 are referred to herein as *paths*. The path is a sequence of positions of a cursor in the window space 10 which correspond to the movement of the wand, thus scheduling the state changes in the system. The cursor position in the window space 10 then maps to a point in a phase space 14 through a callback function. A path through a phase space 14 is a mapping from some time interval  $[0; t_{\text{max}}]$  to the phase space 14. This map need not be bijective or continuous; a path can cross itself, or make abrupt jumps. The path is stored in the phase space 14, not in the window space 10. Thus, a sequence of points of the path is defined with respect to the high-dimensional manifold, and its projection is defined with respect to the particular window space 10 being used.

A path is initially recorded as a set of  $(n+1)$ -tuples, points in the Cartesian product of the  $n$ -dimensional phase space and one-dimensional time. This raw data is smoothed prior to being stored as a C++ path object. The smoothing is done by approximating the original path through this  $(n+1)$ -space with a sequence of spline curves. These splines are also in time as well as in "spatial" dimensions, and are computed in the high-dimensional space. This smoothing is done with a GA, where the bit vector representation of a sequence of spline segments is preferably a vector of fixed-point control points and the fitness function approximates a least-squares error measure integrated over the original path.

The path is drawn through a window space 10 and encoded in a phase space 14. One of the preferred attributes of paths is a record/retrieval functionality which stores and returns temporal information as well as positional information from the user's activities.

This enables users to reflect and revisit the previous decisions or movements in a time critical manner.

### SURFACES AND FIDUCIAL POINTS

On a workstation, a desktop mouse (not shown) is inherently incompatible with  
5 three-dimensional control. Surfaces are instead drawn in the window space 10 and constrain the cursor to the surface, thus compromising with the locally two-dimensional behavior of the mouse. A three dimensional window space surface 40 is shown in Figure 4. Paths can then be recorded on the surface 40 by gestures in two dimensions. The concept of *surface* 40 is also useful in a 3D environment to provide regional differentiation with explicit  
10 geographical representation of subsets of control space.

In order to create surfaces, the window space 10 is investigated and an initial *fiducial point* is formed which can be linked to other fiducial points in the window space 10 by generating a surface 40 through the points, creating a terrain of control values. The terrain may be altered in shape by trimming edges, changing the position of a fiducial point  
15 or adding or removing points. Currently, a plane to displacements is constrained in the direction perpendicular to its surface, this being the most direct conversion of the control region of a 2D mouse to three spatial coordinates. As those skilled in the art will appreciate, more complex surfaces containing folds or intersections may be incorporated as control surfaces by enhancing the 2D navigation constraints of the mouse (not shown), for example  
20 with keyboard (not shown) commands for navigating in a third dimension. However, hybrid interface solutions that require interrupted movements in hand and arm gestures are cumbersome and intrusive to a user's concentration.

Several methods may be employed for creating a smooth surface among a set of fiducial points. Splines are familiar tools in the graphics industry for constructing desired  
25 curved lines or surfaces. One drawback to their application for manifold surfaces is their potential need for multiple control points for assigning smoothing coefficients to a curve. By adding control points, the one-to-one relation is lost between fiducial points and surface-defining points. To preserve this relation, an averaging-displacement method is used for arriving at a smooth surface 40, passing through all the fiducial points and requiring no other  
30 surface-defining points. Beginning with a set of fiducial points defined by two planar

dimensions and one perpendicular dimension, the perpendicular displacement value of each fiducial point is applied at regular intervals across the planar surface. The displacement at each interval across the surface is determined by weighting the displacement according to the distance from each fiducial point to a surface location, and averaging all of the weighted displacements for that location. This procedure is repeated at regular intervals across the surface. This procedure is not applicable if two fiducial points are permitted to share the same planar coordinates with different displacement values.

The architecture of the MC control flow can be summarized as a double-loop feedback and rendering cycle. A flow chart of the MC control flow is provided in Figure 5. As shown in Figure 5, gestures 50 are communicated to the MC 52, which are in turn coupled to functions that generate the window space 54, phase space 56 and paths 58. From the phase space function 54, placement or location information is coupled to both a composition engine 60 and synthesis engine 62. Outputs from both engines 60, 62 are communicated to a sound rendering function 64. Outputs from the paths 58 and window space functions 54 are likewise coupled to a graphics engine 66, the output of which is communicated to an image rendering function 68. Both renderings, sound and image 64, 66 are then fed back to the gesture input device 50.

## EXPLORATION MODES

The model of an observer in exploration modes with the MC is a model of a performing listener. Listening to sounds generated by her or his preceding movements, an observer's cognitive responses are in turn reflected in her or his next movements. Thus, each movement is an articulation of decision-making based upon the evaluation of a preceding acoustic consequence and a proposition for the next.

Navigating a high-dimensional manifold with the MC provides an alternative approach to conventional scientific investigations where all but one dimension is held constant while the remaining dimension is varied. An alternative paradigm for exploration modes in a high-dimensional manifold is in several ways akin to learning a musical instrument:

1. Nonlinearity. The interfaces such as a keyboard on a piano, or valves and fingerholes of wind or brass instruments, have nonlinear relationships to the vibratory

system states as well as the acoustic results, yet one can learn to perform control gestures to extract desired results.

2. Intuitive Orientation. Explorers do not need to attend in detail to the dimensions being varied and how, since this information is encoded by the window space embedding prior to exploration. Explorers can concentrate on grasping an intuitive orientation with respect to the control space.

3. Applicability for Unpredictable Skills. Musical instruments are available for those whose skills vary from novice to virtuoso. A virtuoso is an expert of an instrument by an understanding of its physical properties. She or he knows how to enter motion control to the system in order to achieve desired sounds as well as how to apply acquired listening skills to continuously diagnose the states of the system.

4. Global Orientation. It is worthwhile to note, when observing novice performers' learning processes, that it is more efficient for them to learn an instrument by grasping its whole physical space rather than trying to gather a performance sense by investigating one key or one type of blow stroke at a time. After this global orientation, there will be time for refining individual movements in relation to particular keys or strings for extracting desired tone quality. An easy scalability of control parameter space enables explorers to choose their own orientation scope until they acquire the ability to rapidly fine-tune relations among control variables to achieve desired system states.

#### 20 *Orientation, experiment, and production*

The maturity stage of an observer's interactivity with the system can be described by three stages: orientation stage, experiment stage, and production stage. Each stage has its heuristic value and an observer gains an insight and understanding of the systems while stepping through the stages. Descriptions of these stages are based upon data collected, and are suggestive to adopt alternative and creative ways of exploring computational models.

During *orientation stage*, explorers investigate the whole control space by assigning attributes to the axes of the window space 10. Finding a region of interest, the user refines the scope of the window space 10 by specifying minimum and maximum boundary values of the attributes. Once a good scope of window space 10 is decided an observer can

experiment with the details of the space by choosing generating points, by specifying surfaces and fiducial points and by encoding paths.

5 In the *experimental stage*, explorers learn the temporal sensitivity in state changes of the systems with respect to the sense of speed of their own motions as well as the spatial sensitivity affected by resolution according to the size of the grid in control space. Having found acoustically relevant regions and paths at this exploratory stage, the paths can then be subjected to rigorous experimentation. Gaining a certain degree of familiarity, one can pursue unusual tasks for intermediate experiments; displacements can be performed on source paths by applying transformations such as translation, rotation, augmentation, and  
10 diminution. Translation and rotation affect the values of parameters, not the temporal content of the source path. Augmentation and diminution will affect temporal content as well as parameter values, altering either the rate of change or the duration.

Quick and systematic generation of displacements can be performed in two ways. Out-of-time displacements can be achieved by applying transformation rules to the  
15 source paths to generate batches of files in non-real time. The results are available for real-time review through window space 10. In-time displacements are generated by real-time encoding along with the source path. While a source path playback is initiated as an accompanied event, one can detach the cursor from the path and use it to send additional control messages to generate a variation to the original. This is analogical to the way a jazz  
20 musician generates material in jazz performances. Only in jazz one cannot backtrack whereas with this system and method the user can backtrack all the sources and origins and their relations.

For other examples of unusual tasks, the presently preferred MC provides functionality to bundle several paths and initiate them simultaneously so that an observer  
25 may experience polyphonic retrieval of her or his previous gestures. During this retrieval one may also record yet another path and study the acoustic deviations. By the time an observer steps through all these stages she or he is an expert of the window space 10 and ready to go to *production stage* where she or he decides what data and paths to keep or to discard, documents them, and scripts them as desired. These are subjected to further  
30 refinements and analysis.

Where out-of-time displacements offer a systematic approach to generate



variations and real-time reviews, in-time displacements offer a large variety of playfulness. The latter case is as informative as the former. For example, an explorer can start a second path in conjunction with a source path, and apply displacements as time passes to observe the differences of the two paths in duets while controlling the degree of deviation. This would be a powerful way to generate a pair or a group of modifications with intended degrees of deviation since human ears are good at evaluating fine details of deviations and variations.

## APPLICATIONS

It is important to note that the Manifold Interface is most often applied to modify computational models in real time, which means that an observer while making movements is able to immediately observe the consequences of the movements. This real-time feedback enhances the effectiveness of making motions in a window space 10, because one can allow one's motions to be guided by visual or auditory feedback on a very short time scale. The manifold interface may also be applied in non-real-time feedback systems, which are less efficient due to a time delay between an input and a feedback.

Three applications are presently envisioned for the MC: a physically-based model, a simulated resonance, and an algorithmic musical pattern generation. These applications are described in greater detail below.

### *Multi-dimensional bifurcation scenarios in a simulated chaotic circuit*

Traditional studies of chaotic systems observe bifurcation scenarios by identifying a single bifurcation parameter and varying its value for generating changes in a state of a chaotic system. A Chua's circuit (not shown) belongs to the class of simplest electronic circuits that can generate chaotic signals, and is one of the few known experimental chaotic systems which can be modeled numerically and in computer simulations. Following preliminary experiments with a Chua's circuit for observing acoustic properties of attractors, it was observed that many state producing interesting sounds cannot be achieved by the variation of a single parameter. Using a numerical emulation of the Chua's circuit implemented as a real-time oscillator in the NCSA Sound Server, a multiple-parameter variation technique can be applied from the MC to continuously vary the voltage

values of simulated circuit components. The resulting trajectories of parameter values generate bifurcation scenarios producing acoustic signals that are informative concerning the state of the circuit and are potentially interesting for musical composition. The MC may also be applied to an experimental voltage controlled Chua's circuit for generating composed sequences of state and bifurcations to produce signals for real-time musical performance.

### *Dynamically Controlling Vowel Synthesis*

CHANT, a software-based audio signal generator available from Ircam in Paris, France, synthesizes sound from a description of frequency spectrum characteristics and a simulation of the output of an excitor-resonator system. CHANT waveforms require the specification of seven parameters for each formant in the spectrum. For best result the spectrum should vary over time. With the CHANT libraries installed in the NCSA Sound Server, the manifold interface can generate CHANT sounds in real time. To define a window space 10, specific sounds are associated with specific locations -- generating points -- in the window space 10. Configuring a window space 10 for rendering a CHANT waveform requires four steps:

1. Identify sets of formant parameter values for specific vowel sounds.
2. For each vowel, associate its formant parameter net with unique 3D position in a window space, creating a generating point.
3. Compute the embedding such that all points in the window space have acoustic properties consistent with those of the generating points (smooth transitions occur between generating points).
4. For the examples, shown in Figure 6, create a path 30 in the window space 10 that visits each generating point 60.

For these examples, three formants were rendered, requiring 21 parameters. The parameter changes are nonlinear with respect to one another along the control path. For each generating point 60 eight parameters were defined: the center frequency and bandwidth of the first formant, and the center frequency, bandwidth and amplitude of formants two and three. Four generating points 60 were created; each was assigned a unique vowel sound (/u/, /i/, /c/, or /a:/) and each vowel point was positioned at a unique corner in the window space 10. Amplitude was measured in decibels and center frequency and bandwidth in hertz.

Using the same points as path control points, a path 30 was created passing once through each of the vowels. Signals from five locations 62 on this path 30 are presented in Figure 6. Intermediate positions on the path 30 produce intermediate vowel sounds, such as /U/, which occurs in a location toward the center of the window space 10. In Figure 6, the cursor 64 on the floor is positioned so that its vertical axis intersects the path at the point of the intermediate vowel, /U/.

### *Transformation of Musical Patterns*

Previous examples control simulated signal generators for producing complex tones. The control can be applied in larger scale to signals structured of a series of discrete acoustic events. Messages such as music or speech organized in streams which are parsed into phrases, words, motives, and notes, are a class of signals to which humans devote most of our listening attention in daily life. Composed patterns provide an auditory interface with two desirable features: (1) an acoustic background helping listeners to make comparisons among auditory messages, and (2) a balance of repetition and change helping listeners to remain attentive to significant changes without tiring their ear by attempting to give equal attention to every detail. In a preferred interface prototype, positions in 3D space are associated with transformation of musical materials determining rhythm, pitch and instrument according to positions on three axes. With the preferred MC, musical patterns from the 3D location interface can be further differentiated into the six axes: methodic pitch sequence, melodic rhythm sequence, melodic instrument, tempo, accompaniment, harmonic sequence, accompaniment metric pattern. These elements are independently transformed and combined to create unique patterns at regularly-spaced intersections in the six-dimensional space.

In the presently preferred embodiment of the system and method, the high-dimensional phase space is made up of parameters for controlling sound synthesis. However, as mentioned above this system and method is not limited to controlling sound synthesis. The system and method described herein can be applied to control any high-D computational model. Those models that provide real-time feedback to an observer, are particularly good candidates for application of the manifold interface. Some examples of alternative applications are:

1. the control of the movements of an animated figure;
2. the control of robot motion paths; and
3. the control of states in a description of a medium, such as a particle system or a description of a liquid or gaseous field or volume.

5           The control of the movements of animated figures is important because there are currently a number of applications for controlling animated puppets in real-time. Most of these methods are cumbersome and inefficient, for example, requiring the controlling performed to wear a wired body suit (not shown) because the motion of each parameter of the puppet (not shown) is thought to require a unique body motion from the performer. The  
10 manifold interface can provide needed access to high-dimensional space more efficiently.

          In the control of robot motion paths the computational model includes a digital-to-analog conversion of the high-D signal, applied to a hardware device such as a robotic arm (not shown). Thus, the manifold interface can be extended to control physical devices that can be described as having a smooth high-dimensional phase space 14.

15           In the control of states of a medium, a particle system modeled numerically in the computer and displaying computer graphics of particles has been demonstrated where the particles move and collide. The numerical parameters of the space and the particles are controlled using the manifold interface. In real-time, the motions and interactions of the particles (not shown) change when positions are changed at the interface. Parameters  
20 controlled include gravity, restitution, mass, and friction, for the particles and for the space and its bounding surfaces.

          There are many occurrences of high-dimensional spaces in computational science and engineering. When exploring or navigating such a space, it is inefficient to visit every unique point in the high-dimensional space. Usually there are some subregions of the  
25 space that are of interest, and other subregions that are not. The manifold interface technology hastens the process of identifying the relevant subregions.

          The generating points describe points in a desired High-D subspace, and desired positions for those points in window space 10. The GA fills out a smooth subspace in High-D between these reference points, and provides a mapping to window space 10 from  
30 each point in the High-D subspace. Intuitive 2D or 3D movement can then be used to search this subspace from the window space 10.

5 Freed from complicated control tasks, explorers can concentrate on observing system behaviors as well as their interactivity with system. Control paths provide a form of gesture-based notation. These can be treated as virtual control signals, exported from the manifold interface and stored in files. Thus, the concept of human movement as an exploration of a system can be formalized in a data structure. A capability to encode and formalize control signals suggests a study of control signals with respect to the systems in which they are generated. Such a study may result in further definition of the human inquiry process in an interactive interface.

10 It is to be understood that a wide range of changes and modifications to the embodiments described above will be apparent to those skilled in the art, and are also contemplated. It is therefore intended that the foregoing detailed description be regarded as illustrative rather than limiting, and that it be understood that it is the following claims, including all equivalents, that are intended to define the spirit and scope of the invention.

## WHAT IS CLAIMED IS:

1. A system for interfacing sound synthesis to physical movement, comprising:
  - an interactive sound composition interface (52), the interface coupled to  
5 receive input reflecting movement of a composer;
  - means (56) for generating a phase space (14), the phase space representing at least one dimension of sound in relation to a location of the composer, the location being received from the interactive sound composition interface;
  - means (54) for generating a window space (10), the window space for  
10 translating information from the phase space to render an image of the relation between the location of the composer and sound, the image capable of being displayed to the composer; and
  - a sound synthesizer (64), the sound synthesizer coupled to the interactive sound composition interface to generate sound in response to the location of the composer.
- 15 2. The system defined in claim 1, further comprising a three-dimensional virtual reality environment (10), the three-dimensional virtual reality environment operative for capturing the movement of the composer and coupled to communicate the movement to the interactive sound composition interface.
- 20 3. A method for interfacing sound synthesis to physical movement, comprising the steps of:
  - receiving input (30) reflecting movement of a composer;
  - generating (56) a phase space representing at least one dimension of sound in relation to a location of the composer;
  - translating (66) information from the phase space to render (68) an image of  
25 the relation between the location of the composer and sound;
  - displaying (50) the rendered image to the composer; and
  - synthesizing sound (64) in response to the location of the composer.

4. The method defined in claim 3, further comprising the step of generating a window space (10) for displaying the rendered image to the composer.

5. The method defined in claim 3, further comprising the step of providing an interactive sound composition interface, the interface coupled to receive the input reflecting movement of the composer.

6. The method defined in claim 3, further comprising the step of providing a three-dimensional virtual reality environment operative for capturing the movement of the composer.

7. The method defined in claim 6, further comprising the step of communicating the captured composer movement to the interactive sound composition interface.

8. A system for navigating a manifold in a high-dimensional space, comprising:

a three-dimensional virtual reality environment (10) operative to capture movement of an object relative to the manifold and coupled to communicate the captured movement;

means (56) for generating a phase space (14) representing at least one dimension of the manifold in relation to location of the object, the location being received from the three-dimensional virtual reality environment; and

means (54) for generating a window space, the window space for translating information from the phase space to render an image of the relation between the location of the object and the manifold, the image capable of being displayed.

9. The system defined in claim 8, wherein the manifold is sound generated.

10. The system defined in claim 8, wherein the object comprises a robot.

11. The system defined in claim 8, wherein the object comprises an animated figure.

12. A method for navigating a manifold in a high-dimensional space, comprising the steps of:

5 capturing movement (30) of an object in relation to the manifold in a three-dimensional virtual reality environment;

communicating the captured movement;

generating (50) a phase space (14) representing at least one dimension of the manifold in relation to location of the object, the location being received from the three-  
10 dimensional virtual reality environment; and

generating (54) a window space (10) for translating information from the phase space to render an image of the relation between the location of the object and the manifold, the image capable of being displayed.

13. The method defined in claim 12, further comprising the step of storing  
15 and retrieving said captured movement to reproduce a sequence.

14. The method defined in claim 12, further comprising the step of performing transformations upon captured movements for further exploration and differentiation of said space.



1/7

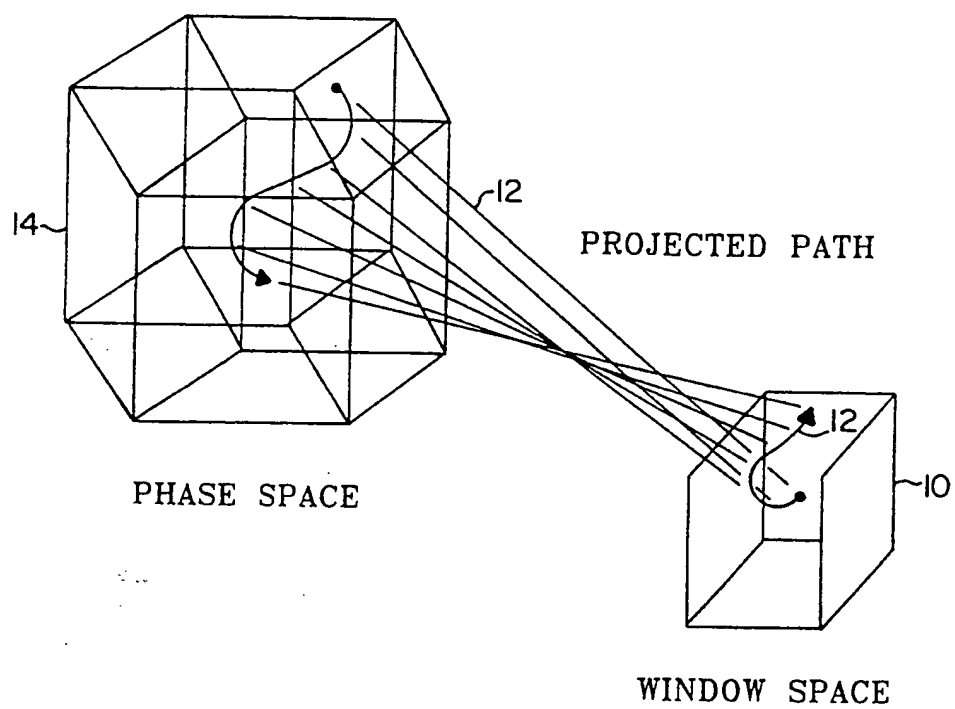


Fig. 1

2/7

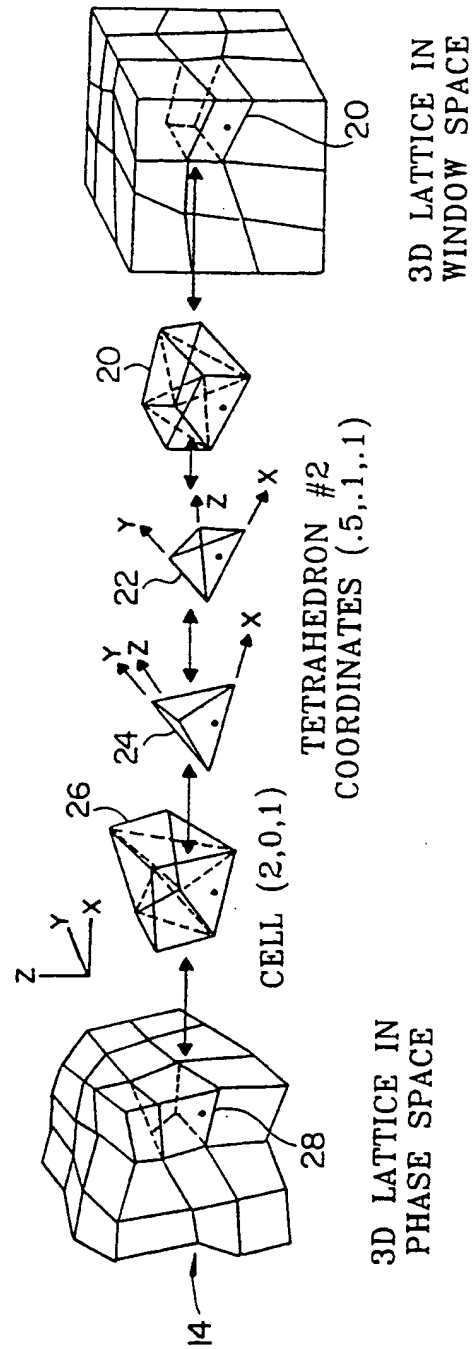


Fig. 2

3/7

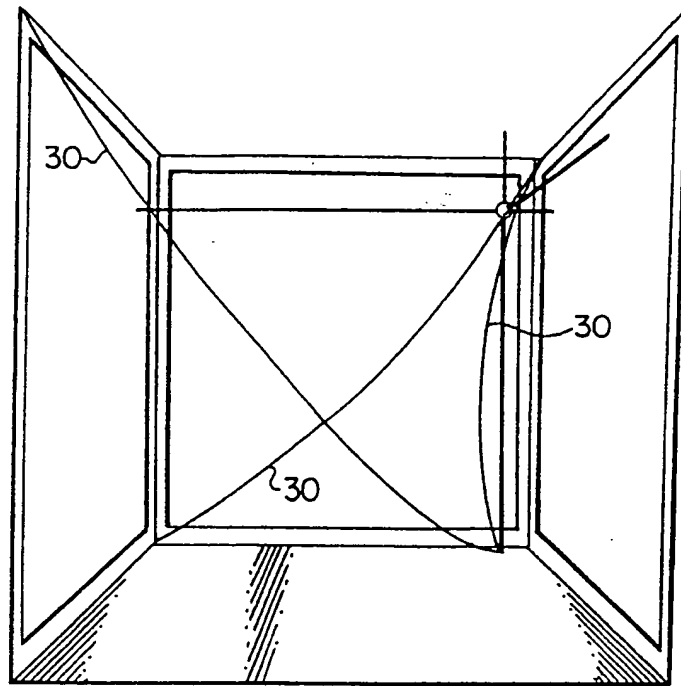


Fig. 3

4/7

BEST AVAILABLE COPY

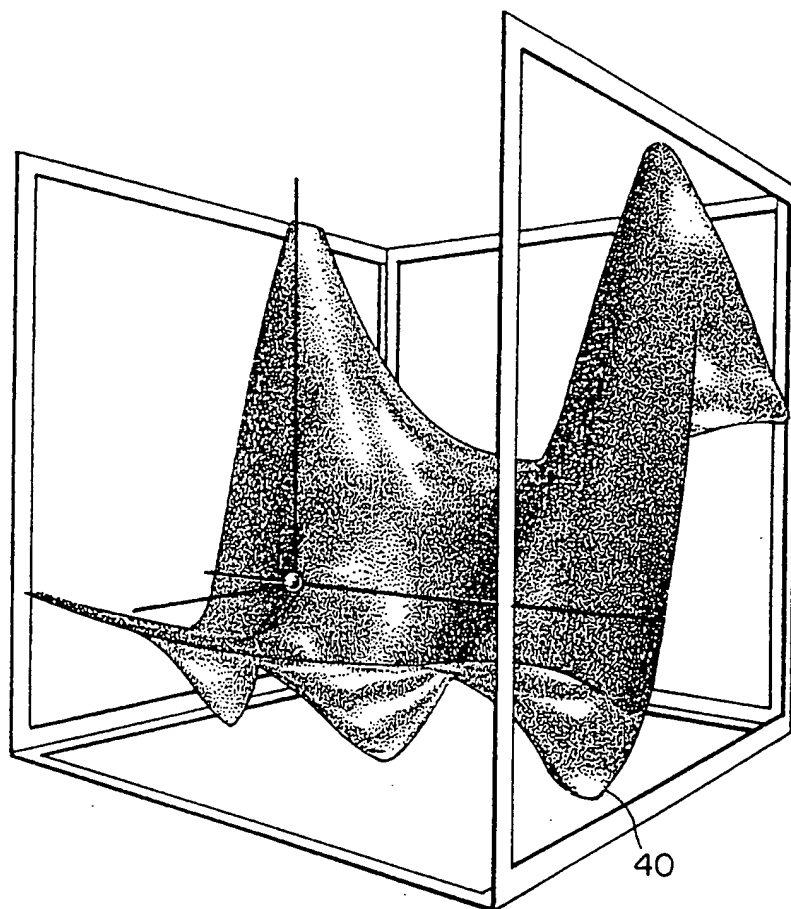


Fig. 4

5/7

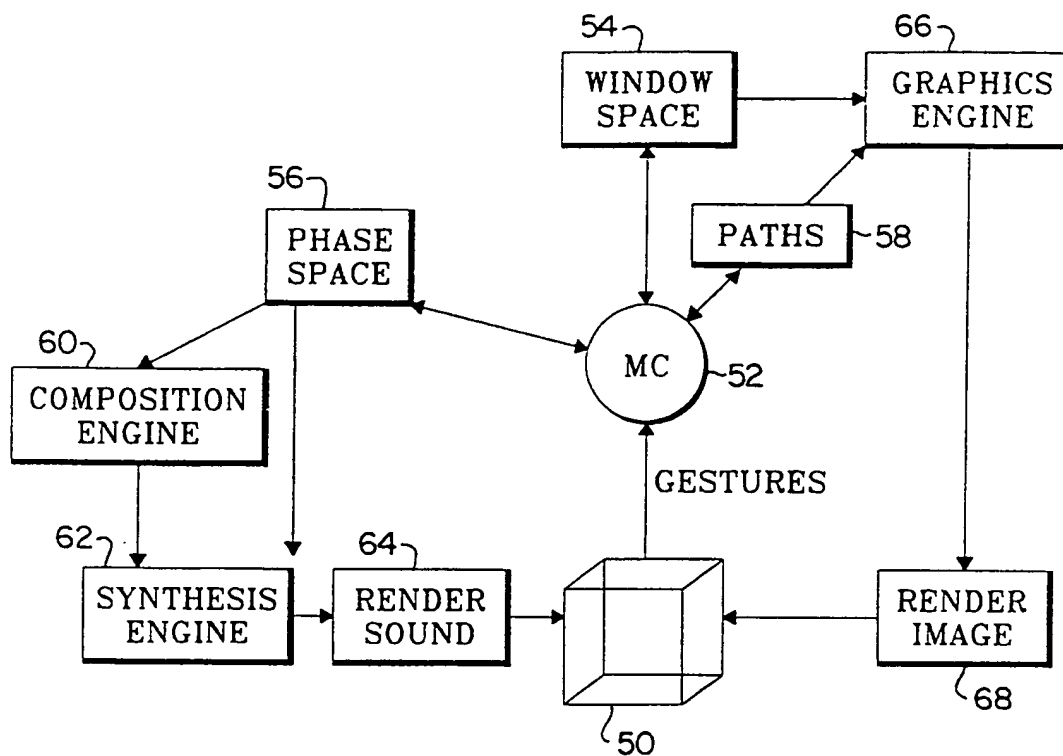


Fig. 5

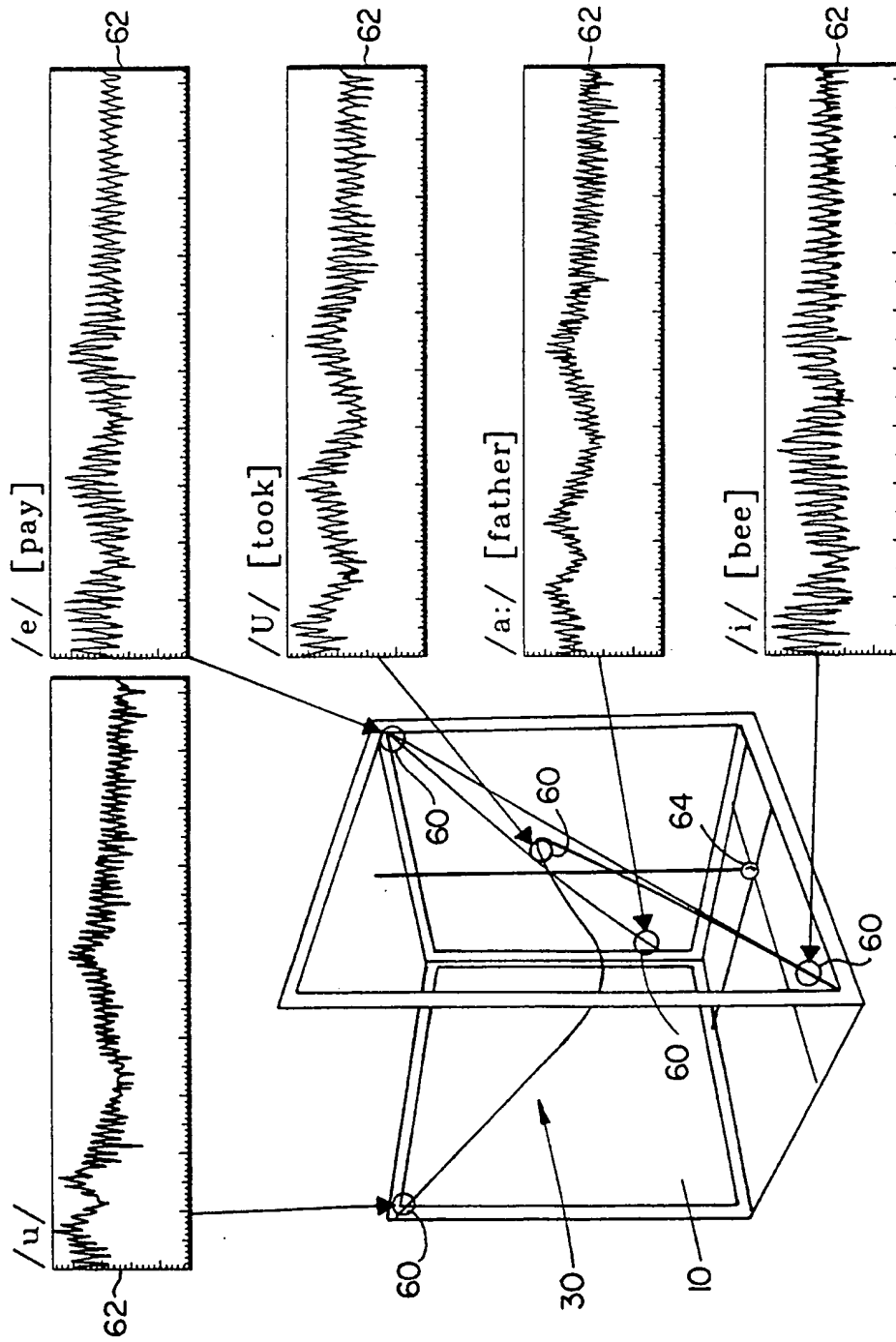


Fig. 6

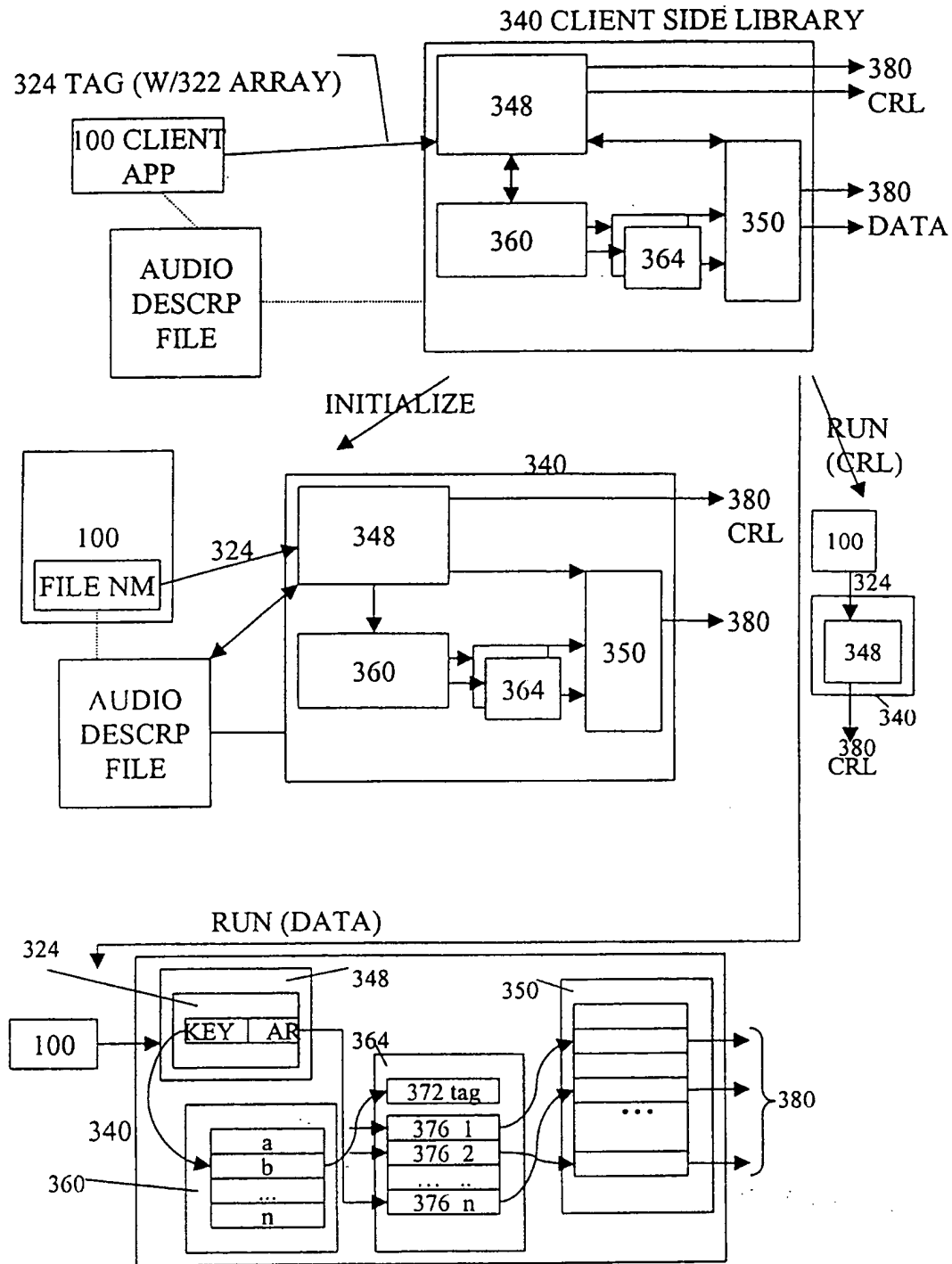


Fig. 7

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US98/18290

**A. CLASSIFICATION OF SUBJECT MATTER**

IPC(6) : G10L 5/02

US CL : 704/258; 364/578; 381/17

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 704/258, 200, 270, 276; 364/578; 381/17,18

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, DIALOG, Internet

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CHOI et al. Interfacing sound synthesis to movement for exploring high-dimensional systems in a virtual environment. Systems, Man and Cybernetics, 1995. Intelligent Systems for the 21st Century., IEEE International Conference on. 22-25 October 1995, Vol 3, pages 2772-2777, entire reference.	1-14
X	BARGAR et al. Model-based interactive sound for an immersive virtual environment. Proceedings of the International Computer Music Conference, 1994. September 1994, pages 471-477, entire reference.	1-14

☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:	*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
*A* document defining the general state of the art which is not considered to be of particular relevance	*X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
*E* earlier document published on or after the international filing date	*Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
*L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*A* document member of the same patent family
*O* document referring to an oral disclosure, use, exhibition or other means	
*P* document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

07 JANUARY 1999

Date of mailing of the international search report

21 JAN 1999

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

David R. Hudspeth

Telephone No. (703) 308-4825

Jon Hill



## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US98/18290

## C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CHOI et al. Human-machine performance configuratin for computational cybernetics. Systems, Man and Cybernetics, 1997. Computatiaonal Cybernetics and Simulation., IEEE Internation Conference on. 12-15 October 1997, Vol. 5, pages 4254-4259.	1-14
A	CRUZ-NEIRA et al. The cave audio visual experience automatic virtual environment. Communications of the ACM. June 1992, Vol. 35, number 6, pages 65-72.	1-14
A	AXEN et al. Investigating geometric data with sound. Proceedings of International Conference on Auditory Display. 4-6 November 1996.	1-14
A	US 5,513,130 A (REDMOND) 30 April 1996 (30.04.96).	1-14
A	US 5,587,936 A (LEVITT ET AL.) 24 December 1996 (24.12.96).	1-14